A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes)

Verkuil, Yvonne I.; Piersma, Theunis; Baker, Allan J.

Published in:
Molecular Phylogenetics and Evolution

DOI:
10.1016/j.ympev.2010.06.010

Publication date:
2010

Molecular Phylogenetics and Evolution 57 (2010) 411–416

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes)

Yvonne I. Verkuil [a,b,*], Theunis Piersma [a,c], Allan J. Baker [b,d]

[a] Animal Ecology Group, Centre of Ecological and Evolutionary Studies (CEES), University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands
[b] Department of Natural History, Royal Ontario Museum, 100 Queen's Park Crescent, Toronto, ON, Canada M5S 2C6
[c] Department of Marine Ecology, Royal Netherlands Institute for Sea Research (NIOZ), P.O. Box 59, 1790 AB Den Burg, Texel, The Netherlands
[d] Department of Ecology and Evolutionary Biology, University of Toronto, 25 Harbord Street, Toronto, ON, Canada M5S 3G5

ARTICLE INFO

Article history:
Received 25 March 2010
Revised 13 June 2010
Accepted 15 June 2010
Available online 19 June 2010

Keywords:
mtDNA
Control region
Gene duplication
Concerted evolution
Philomachus pugnax
Sandpiper
Aves

ABSTRACT

Although the mitochondrial genome in birds has highly conserved features, with protein genes similar to mammals and amphibians, several variations in gene order around the hypervariable control region have been found. Here we report a novel gene arrangement around the control region in shorebirds (Charadriiformes). In ruffs *Philomachus pugnax,* the mitochondrial genome between cytochrome *b* and 12SrRNA was over 1.5 kb longer than reported for other Charadriiformes and contained a duplication of the control region together with NADH dehydrogenase subunit 6 (ND6) and the adjacent transfer RNAs: $tRNA^{Pro}$ and $tRNA^{Glu}$. Sequence data from 68 individuals from several stopover and breeding populations show that the duplication is widespread in ruffs. Similar gene re-arrangements have been found independently in unrelated tube-nosed seabirds and spoonbills.

1. Introduction

Two hundred and fifty million years ago, sometime after the bird-crocodile split but before the radiation of modern birds, a novel gene re-arrangement near the control region (CR) evolved in the avian mitochondrial genome (Desjardins and Morais, 1990, 1991; Quinn and Wilson, 1993). Since then gene duplications and re-arrangements have arisen independently several times among birds (Gibb et al., 2007; Haring et al., 1999; Mindell et al., 1998; Singh et al., 2008). They occur more often around the CR, a site for initiation and termination of DNA replication, than elsewhere in the mitochondrial genome (Fujita et al., 2007). The duplication of the CR often involves changes in the gene order and duplications of the protein-coding NADH dehydrogenase subunit 6 (ND6) and the flanking transfer RNAs (Pereira (2000); Fig. 1).

Sequence variation in the highly polymorphic CR and variations in gene order have both been applied as markers in phylogeographic and population genetic studies (Avise, 2000; Bensch and Härlid, 2000; Gibb et al., 2007; Mindell et al., 1998; Singh et al., 2008). However, gene duplications and re-arrangements complicate the usage of the CR as a marker. Co-amplification of the duplicate copies can result in sequences with "apparent heteroplasmy". Also CR sequences obtained across individuals, populations or species may not be homologs (i.e., one CR may be preferentially amplified and sequenced for some samples, while the second copy may be obtained for others, resulting in comparisons across non-homologous sequences). In this case observed sequence variation in the CR can be either paralogous (comparisons among gene duplicates) or orthologous (inter-individual comparison). This is a serious issue since it does not present any obvious symptoms such as "apparent heteroplasmy."

This paper reports on gene duplications and sequence similarity in the mitochondrial (mtDNA) of a basal Charadriiformes species (Baker et al., 2007), the ruff (*Philomachus pugnax*, Scolopacidae). Ruffs are an interesting case for studies of the relationship between phenotypic and genotypic variation in populations as they appear genetically variable (Segre et al., 1970) though geographically unstructured (Verkuil, 2010). Morphometric variation, however, indicates segregating selective pressures in areas within its vast migratory and breeding range (Karlionova et al., 2007). Sequence variation in the hypervariable CR has been shown to be informative in determining genetic population structure in shorebirds (Buehler and Baker, 2005; Buehler et al., 2006; Wenink et al., 1993, 1994). In ruffs, however, sequences of the CR obtained with primers designed in the conserved regions of the CR and ND6, indicated multiple copies. This hinted at re-arrangements in the gene order around the CR, although other potential sources of duplicates are

* Corresponding author at: Animal Ecology Group, Centre of Ecological and Evolutionary Studies (CEES), University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands. Fax: +1 416 586 8022.
E-mail address: yvonne_verkuil@hotmail.com (Y.I. Verkuil).

1055-7903/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2010.06.010

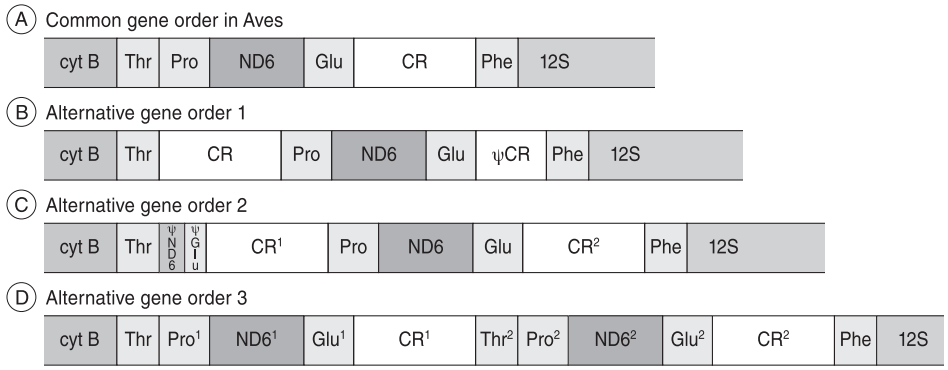

Fig. 1. Gene arrangements around the control region (CR) in birds: (A) gene order reported as common in Aves (Desjardins and Morais, 1990); (B) first alternative gene order found in Aves (Mindell et al., 1998; Bensch and Härlid, 2000); (C) gene order in parrots (Eberhard et al., 2001), and (D) gene order in albatrosses (Abbott et al., 2005). Ψ; locus is degenerate.

the amplification of nuclear mitochondrial pseudogenes (numts) along with the target mtDNA sequence or heteroplasmy, i.e. the coexistence of multiple copies of mtDNA molecules within individuals.

Duplicates are often degenerate but also can have high sequence similarity, as CR copies may be kept similar by concerted evolution, which may act a repair mechanism to avoid transcription error (Tatarenkov and Avise, 2007). In vertebrates, evidence for concerted evolution of duplicate CR regions has been found in fish (Tatarenkov and Avise, 2007), frogs (Kurabayashi et al., 2008), snakes (Jiang et al., 2007), and birds (parrots (Eberhard et al., 2001), albatrosses (Abbott et al., 2005) and warblers (Singh et al., 2008)). We discuss two alternative explanations for the high similarity of the two CRs in ruffs: (1) recent duplication of the CR with enough time for fixation but not enough to accumulate mutations, or (2) concerted evolution (Tatarenkov and Avise, 2007).

2. Methods and material

Blood samples of ruffs (*Philomachus pugnax*) were collected on spring staging sites in The Netherlands (by J.C.E.W. Hooijmeijer and authors), Belarus (by N. Karlionova and P. Pinchuk) and in breeding areas in Sweden (by K.A. Thuman), Finland (by D.B. Lank) and Siberia, Russia (by M.Y. Soloviev, P.S. Tomkovich et al.,). Samples were stored in 98% ethanol at -80 °C. All except the Swedish samples are curated in the collections of the Royal Ontario Museum, Toronto, ON, Canada, and the University of Groningen *Shorebird LifeLines Blood Bank*, The Netherlands.

DNA was isolated with DNeasy Blood and Tissue Kit (Qiagen) or by standard phenol–chloroform extractions and stored at −20 °C. To screen for genetic variation throughout the migratory and breeding range of ruffs, the 5′-end of the CR was amplified using shorebird primers anchored in ND6 and the conserved blocks in the CR: ND6L3 & H451/H772 (Wenink et al. (1994); see Table 1). However, the sequences obtained varied in size which indicated multiple copies of the CR. Therefore, to explore the gene order around the CR and to exclude the possibility of amplifying nuclear copies, long template (LT) was amplified that spanned the region between cytochrome *b* and the 12 S ribosomal RNA genes (12S), using Expand Long Template PCR (Roche, Basel, Switzerland). Subsequently the LT amplicon was used as a template to obtain partial sequence of the region. The LT primers used were b86 with H1827 for ruff 1479281 (caught on 20 December 2004 in The Netherlands), b3 with 16ScR for ruff 1481191 (caught on 4 March 2005 in The Netherlands, and resighted in Russia on 15 May 2005, Surgut, Khanty/Mansi, 60°10′N, 74°00′E), and b71 with H1390 for both individuals (Table 1). The LT PCR profile was 2 min denaturation at 94 °C, followed by 10 cycles of 92 °C for 30 s, 63 °C for 30 s and 68 °C for 12 min and 25 cycles of 92 °C for 30 s, 63 °C for 30 s

Table 1
Primers used for amplification of long template DNA from which nested amplifications were made to establish the gene order around the mitochondrial control region of ruffs (*Philomachus pugnax*). For each primer, the location is given and a short explanation of why it was applied.

Region	Primers	SEQUENCE	Application	Source
Cytochrome *b*	b86	5′-TGAATNGGNAGCCARCCNGTAGAACACCC-3′	Long template-1	O. Haddrath (unpublished)
	b3	5′- GGACGAGGCTTTTACTACGGCTC-3′	Long template-2	T. Burke (unpublished)
	b71	5′- TGGGAGGAGTACTAGCTCTGGCAGCCTC-3′	Long template-3	This study
	b5	5′-TTCCACCCCTACTTCTCACTAAAGA-3′		T. Burke (unpublished)
	cytbend	5′-CACGACGTTGTAAAACGACAGCCAACCYAGAACACCC-3′		T. Paton (unpublished)
$tRNA^{Pro}$	PropR	5′-AATACCAGCTTTGGGAGTTGG-3′	*Backwards* into CR; no microsatellite	This study
NADH6	ND6p2R	5′-TGGTTACTGTTGACAGTGGGGGA-3′	*Backwards* into Domain II-BSB	This study
	ND6p6R	5′ –CTGTTGCAGGATGTGGTTGGCT-3′	*Backwards* into Domain III	This study
	ND6L3	5′-ACTGCTCGAATCGCCCCACGAG-3′	Into C-string	This study
	ND6HR	5′-TYCGNATRGATTTTAGYGGGGT-3′		O. Haddrath (unpublished)
Control region	H451	5′-CCTGAAGCTAGTAACGCAGGAC-3′	Into C-string	This study
	L141	5′-TCCATTAATCTACAACCGGGCT-3′	Into ψ-$tRNA^{Thr}$	This study
	L/H402	5′-TGAAATCAGCAACCCGGCGTAAG-3′	Into ψ-$tRNA^{Thr}$	This study
	L438	5′-TCACGTGAAATCAGCAACCC-3′	*Forward* in 12S; into microsatellite	Wenink et al. (1993)
	L716	5′-ACTTTGGCCCTCAGGCGTTACTG-3′	*Forward* into ND6; no microsatellite	This study
12SrRNA	H1537	5′-TGACCGCGGTGGCTGGCACAA G-′3	Into microsatellite	O. Haddrath (unpublished)
	H1827	5′-GCACCGCCAAGTCCTTAGAG -3′	Long template-1	M. Burbidge (unpublished)
	16ScR	5′-TTCTCCAAGGTCGCCCCAACC-3′	Long template-2	O. Haddrath (unpublished)
	H1390	5′-TCAAGGGCATTTACACTGGGGCGCAGATAC-3′	Long template-3	This study

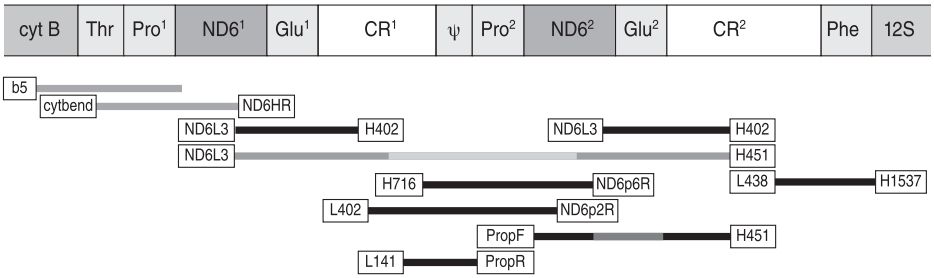

Fig. 2. Overview of the gene order between cytochrome *b* and the 12SrRNA gene in ruffs *Philomachus pugnax*. Amplified fragments (and their relative sizes) used to make the alignment and the primers used in amplifications are given (see also Table 1). Black segments are from nested amplifications off long templates; grey segments indicate sequence from genomic DNA. Dashed parts of PCR products were not sequenced. *Ψ*; degenerate sequence block of 88 nt.

and 68 °C for 12 min + 20 s cycle elongation for each successive cycle. Final elongation at 68 °C was for 7 min.

The amplified LT mtDNA segments were excised from agarose gels, purified by spinning through filter tips and stored at −20 °C. Nested amplifications were performed with primers anchored in cytochrome *b*, 12SrRNA, tRNA^{Pro} and ND6 (see Table 1 and Fig. 2 for primer sequence and location). To isolate the 5′-ends of the CRs, PCR products were amplified off LT (ruff 1479281) and genomic DNA (ruff 1481191) with internal primers ND6L3 and H402, and cloned into pCR®2.1 following the manufacturer's instructions (Invitrogen). Off each ligation, 10 positive clones were sequenced with M13 primers. Only sequence variation that was observed in >1 clone was considered real variation. For comparison, additional PCR products using genomic templates were obtained for two individuals, 1463006 (caught on 17 March 2004 in The Netherlands) and SIB-258 (nestling DNA collected on 17 July 2004 on Taymir Peninsula Siberia, 74°09′N, 99°34′E).

A segment of CR that can only be obtained when the alternative gene arrangement was present was amplified in 68 ruffs, using the forward CR primer L141 and the reverse primer PropR anchored in tRNA^{Pro} (Table 1). Individuals were sampled when breeding in Sweden (n = 20), Finland (n = 10), Russian Arctic (Taymir Peninsula or further east, n = 16) and Norway (n = 2), or were migrating through Belarus (n = 9) or The Netherlands (n = 11).

The amplification profile used for all short-range PCRs was 2 min denaturation at 95 °C, followed by 36 cycles of 94 °C for 45 s, 53–57 °C for 45 s, 72 °C for 1.3 min, followed by a final 7 min elongation at 72 °C. PCR products were gel-purified and sequenced. PCR products were prepared for sequencing using BigDye Terminal Cycle Sequencing reagents according to the manufacturer's instructions (Applied Biosystems, Foster City, USA), and were sequenced on an ABI 3100 automated sequencer.

Sequences were edited and aligned in Chromas Pro 1.33. To manually edit sequences and alignments, MEGA 3.1 (Kumar et al., 2004) was used. Sequences were aligned to homologous sequences of closely related shorebirds, red knot (*Calidris canutus*, GenBank AY198171 (Buehler and Baker, 2003), EF373071, EF373123 (Baker et al., 2007), and ruddy turnstone (*Arenaria interpres*, GenBank NC_003712) (Paton et al., 2002). To control for possible nuclear copies, sequences obtained from genomic DNA were only used when they could be fully aligned with long template sequences. Secondary tRNA structure was simulated in tRNA scan-SE (Lowe and Eddy, 1997). To indicate duplication of genes, the notation given in Fig. 1 is followed.

3. Results

The complete long template (LT) PCR product between the 3′-end of cytochrome *b* and the 5′-end of 12S rRNA was obtained for two individuals. Based on the published mitochondrial gene order and CR size in shorebirds closely related to ruffs (Buehler and Baker, 2003; Paton et al., 2002), the expected size of the LT PCR product (ranging from cytochrome *b* into 12SrRNA gene) was ~3.2 kb. The LT fragment obtained in both individuals was 4.5 kb (primer combination b71 and H1390), and was confirmed with two different primer combinations (Table 1). Clones of products amplified off the LT with internal primers ND6L3 and H402 yielded two different CR sequences in each individual (Fig. 3). The assembly of sequences from nested amplifications off the LT (using primers anchored in cytochrome *b*, 12SrRNA, tRNA^{Pro} and ND6) showed that ruffs have duplicates of the CR and adjacent genes ND6, tRNA-^{Pro}, and tRNA^{Glu} (Fig. 2, GenBank GQ255993). All sequences aligned with sequences of ruddy turnstones and red knots, and had a very similar base composition and the same general molecular organization in three CR domains as other birds (Pereira et al., 2004).

Amplification of Domain I of the CR with primers ND6L3 and H451 off genomic DNA (of ruff 1463006) yielded PCR products of two different sizes: one band of the expected size of about 1 kb and one band of about 2.5 kb. Apparent heteroplasmy was observed in the sequences of the 2.5 kb fragment, indicating that both CRs were included in the 2.5 kb fragment (Fig. 3). Sequences of the 2.5 kb fragment aligned perfectly with the LT clones and with the genomic sequence from SIB-258 from the Siberian breeding population. Within individuals, the sequences showed "apparent heteroplasmy" at nine positions at the 5′-end of Domain I (Fig. 3). Domain I in ruffs was similar to the Domain I of red knots (Fig. 4).

Domains II of the two CRs were identical. Sequences of Domain II obtained from the 1 and 2.5 kb amplicons using the primers L438 and H1537 were identical for both LT and genomic templates. Also, sequences from amplifications off LT, either anchored in tRNA^{Pro} backwards into CR^1 or in 12S backwards in CR^2 (see Fig. 2, Table 1) were in perfect alignment.

Domain III of the two CRs was nearly identical until the TACAT promotor near the 3′-end. After the TACAT promotor the first CR had 88 nt of apparently random sequence, which ended in a complete functional copy of tRNA^{Pro} (Fig. 4). The tRNA^{Pro} at the 3′-end of CR^1 and the tRNA^{Pro} downstream of cytochrome *b* were identical and showed a functional clover leaf folding pattern. The second CR aligned throughout with Domain III of red knots and ruddy turnstones, ending with a 4-nt repeat sequence (microsatellite) of ~40 repeats which was very similar in structure to red knots and ruddy turnstones. Overall, the two CRs were similar, with the first CR being 1120 nt long and the second 1183 nt, as the latter included the 4-nt repeat at the 3′-end.

Two copies of ND6 were found. The first ND (ND6^1) was 540 nt long between tRNA^{Pro} and tRNA^{Glu}. Cloned products of the amplicons of the 3′-end of ND6, obtained with the primers ND6LR3 and H402, were identical. ND6^2 was followed by a complete tRNA-^{Glu}. Both copies aligned with ND6 of ruddy turnstones (Paton et al., 2002). Sequences of the amplicons using primers anchored in

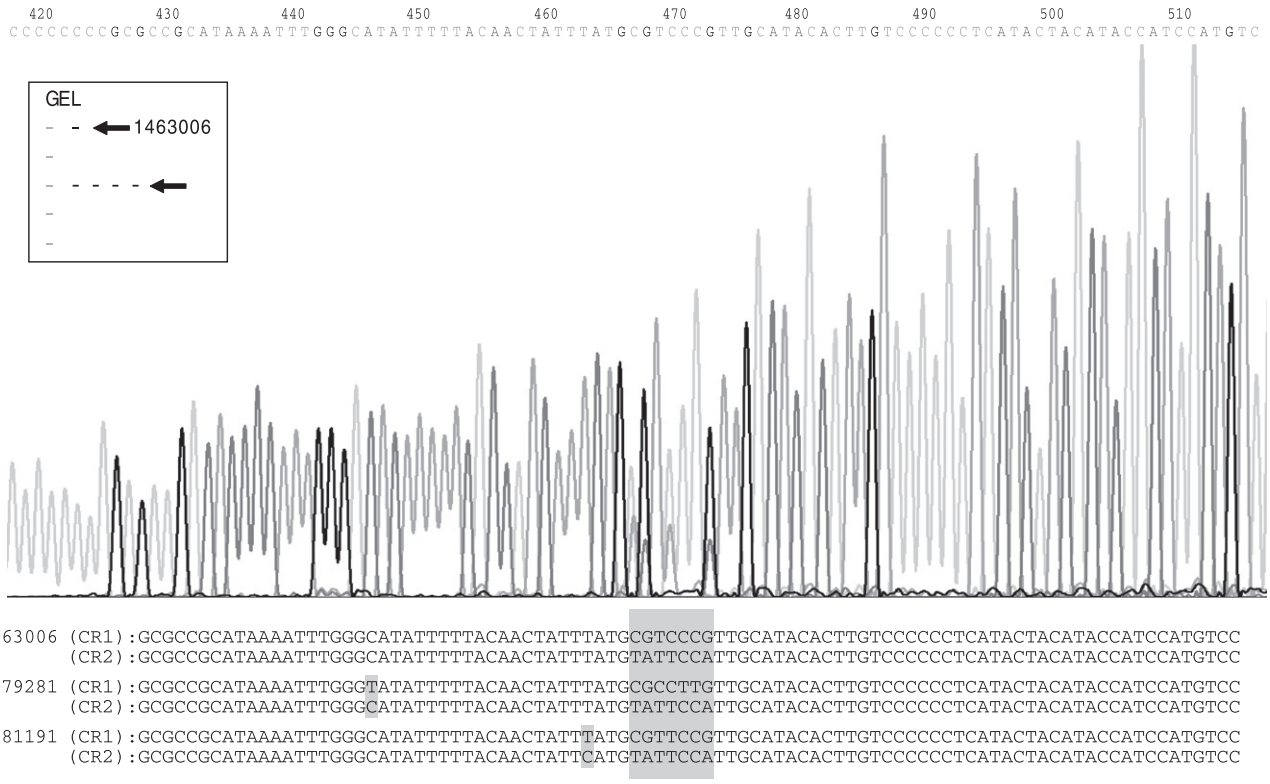

Fig. 3. Apparent heteroplasmy in the control region of ruffs *Philomachus pugnax* due to gene duplication. The chromatogram starts at the C-string at the 5'-end and continues 55 nt into the CR; the double peaks indicated 'apparent heteroplasmy'. This sequence is part of a 2500 nt PCR product obtained from amplification off genomic template of ruff 1463006 with a primer anchored in ND6 and the CR primer H451 (see inset). Below are sequences of clones of amplicons off Long Templates of ruff 1481191 and ruff 1479281 showing that the 'apparent heteroplasmy' is due to two different control regions (CR^1 and CR^2). Inset: graphical presentation of the gel representing amplification in four individuals of which one showed the larger band that indicated the duplication.

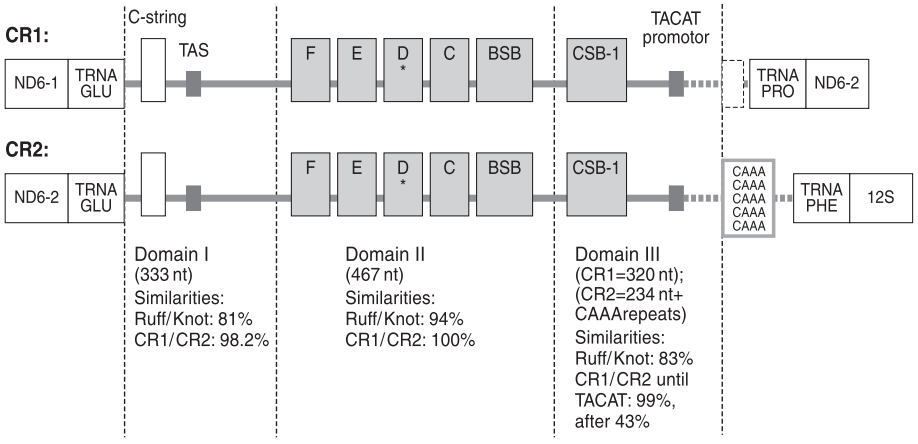

Fig. 4. Schematic representation following Buehler and Baker (2003) of the two control regions (CR) in ruffs *Philomachus pugnax* and comparison with control regions elements in the closely related red knots *Calidris canutus*. Conserved sequence blocks (CSB) and bird similarity box (BSB) are indicated in grey, (*) indicates 1 nt in the conserved sequence blocks where ruffs and red knots were different. The two CRs in ruffs were similar until the TACAT promotor. The dotted box in CR^1 indicates degenerate sequence (88 nt). The striped box indicates the 4-nt repeat at the end of CR^2.

cytochrome *b* and $ND6^1$ or in CR^1 and $ND6^2$ were in perfect alignment (Fig. 2 and Table 1).

To evaluate paralogous and orthologous sequence similarity in Domain I, sequences of the two cloned individuals were compared. Sequence of 386 nt was obtained, and the 333 nt of Domain I of both CRs were aligned (GenBank GQ 174508, 174509, 1874510, 174511). The average sequence divergence of the orthologous copies of the first domains of the two CRs was 1.5% (2.1% in CR^1 and 0.9% in CR^2). Within individuals, average sequence divergence of Domain I was 1.8% (2.4% in ruff 1479281 and 1.2% in 1481191). Most variable sites were located at the 5′-end. Downstream, at position 236, a shared substitution in the two control regions was found within an individual that was not shared by the two control regions in the other individual (Fig. 5).

A large fragment of CR^1 was successfully amplified from 68 ruffs from a variety of breeding and migratory staging sites. The amplicons started at position 124 using forward primer L141 and ended in $tRNA^{Pro}$ using reverse primer PropR, indicating that all individuals had the alternative gene arrangement. The average pairwise sequence divergence of the orthologous copies of Domain I and II

Fig. 5. Sequences of positions 211–272 in Domain I of the control region, indicating a shared substitution (in grey) of the two control regions within an individual that is not shared with the two control regions in the other individual. Sequences were obtained from clones of amplicons off long templates of ruff 1481191 and ruff 1479281. CR^1 and CR^2 are the two different control regions. The highlighted substitution is located at position 236.

between each pair of the 68 individuals was 5.5 ± 0.01% (28 segregating sites in 504 nt). Within the 209 nt fragment that overlapped with Domain I obtained from the clones from individuals the average orthologous pairwise divergence was 7.7 ± 0.02%.

4. Discussion

The mitochondrial genome of the ruff between cytochrome *b* and 12SrRNA was over 1.5 kb longer than reported for other Charadriiformes (Buehler and Baker, 2003; Paton et al., 2002). This additional piece included full duplications of the CR, ND6, $tRNA^{Pro}$ and $tRNA^{Glu}$. Since the duplicate sequences were obtained from a 4.5 kb long template, heteroplasmy in individuals can be excluded as the source of sequences variation. Also, the size of the LT amplicon was larger than most nuclear mitochondrial pseudogenes (numts) reported in birds (Pereira and Baker, 2004). Hence numts are an unlikely source of sequence variation. Moreover, in the ND6 coding genes sequenced from LT, no stop codons or unusual frame shifts were found; hence the LT sequence is unlikely to be a numt. The duplication was present in all 68 samples obtained throughout the species distribution range, indicating that duplicate state of the CRs is not a transient or unstable feature found in a particular individual or deme.

The length of CR^2 (1183 nt) was comparable to CRs in red knots, ruddy turnstones and blackish oystercatchers *Haematopus ater*, which are 1168, 1172 and 1239 nt, respectively (Buehler and Baker, 2003; Paton et al., 2002). CR^1 was shorter because the 4-nt microsatellite-like repeat was absent. Due to the duplications, together with the albatrosses (Abbott et al., 2005) and black-faced spoonbills *Platalea minor* (Cho et al., 2009), ruffs possess the largest control region area (between cytochrome *b* and 12SrRNA) reported for Aves.

The re-arrangement in ruffs reflects a gene order that is unique in the order Charadriiformes, in that ruffs have full duplicates of the CR, ND6, $tRNA^{Pro}$, and $tRNA^{Glu}$. Alternative gene orders in birds often involve pseudogenes and a degenerate copy of the CR (Mindell et al., 1998) as in *Phylloscopus* warblers (Bensch and Härlid, 2000) and in Falconiformes (Cadahia et al., 2009). Gene orders similar to ruffs, with two almost identical copies of the CR, were previously reported for blackcaps (*Sylvia atricapilla*) and reed warblers (*Acrocephalus scirpaceus*) (Singh et al., 2008), but in these species no sequences of the region before the first CR is available so it remains unclear whether they had a duplicate ND6 and tRNA genes. In *Amazona* parrots (Eberhard et al., 2001), osprey (*Pandion haliaetus*) and ivory-billed aracari (*Pteroglossus azara*) (Gibb et al., 2007), the first $tRNA^{Pro}$, ND6 and $tRNA^{Glu}$ are degenerate or pseudogenes. The black-faced spoonbill has a duplicate of the CR and the adjacent cytb, ND6, $tRNA^{Pro}$, and $tRNA^{Glu}$, and similar to ruffs most variation between CRs is found in the 3′-end (Cho et al., 2009). In *Thalassarche* albatrosses (Abbott et al., 2005) the duplicate CR is also flanked by duplicate ND6, $tRNA^{Pro}$, and $tRNA^{Glu}$; the only difference between the gene arrangements in albatrosses and ruffs is that albatrosses had a second copy of $tRNA^{Thr}$. These similarities across taxa must represent convergent evolutionary events in bird genomes (Bensch and Härlid, 2000).

The observed high level of similarity between the two CRs (especialy in Domain II) in ruffs might be a consequence of: (1) recent duplication of the CR, or (2) concerted evolution. In ruffs paralogous sequence divergence was relatively low, being 1.8%. This value is similar to *Amazona* parrots, where the average sequence divergence of the paralogous CR copies was 1.4% and CRs were considered to be under concerted evolution (Eberhard et al., 2001). However, in our study the comparison between orthologous and paralogous sequence variation is hampered by sample size as paralogous copies were obtained for only two individuals. The supposed mechanisms for concerted evolution are gene conversion (sequences stay homogenized by recombination) or tandem replication (sequences are kept homogenized due to replication slippage) (Kurabayashi et al., 2008; Eberhard et al., 2001). Of these mechanisms only gene conversion allows for unevenness in sequences similarity between Domains as was observed in ruffs. The supposedly functional Domain II is most similar between copies (Fig. 4). In this Domain the observation of a shared substitution between the two CRs within individuals but not between individuals (Fig. 5), suggests concerted evolution (but this phenomenon can also be explained by homoplasy). Sites in Domain I provided a situation opposing concerted evolution as substitutions were shared between orthologs but not between paralogs (Fig. 3.)

As the possibility of concerted evolution remains unresolved, we continue to consider the hypothesis of recent fixation of duplicates. Under the scenario of recent fixation, there would not have been enough time to accumulate mutations, so similarity between copies is expected, and, beyond functional constraints, mutation should be randomly distributed. We, however, observed a considerable unevenness in the distribution of substitutions. All variation between CR^1 and CR^2 was found near the 5′-end of Domain I and 3′-end of Domain III. Also, the degeneration of the microsatellite repeat at the 3′-end of Domain III indicates that the duplication was not recent. Alternatively, the duplication event itself created the observed unevenness in distribution of substitutions, given its position at the 5′- and 3′-end of the duplicated segment. The probability of fixation of a gene duplication increases with decreasing effective population size (N_e) as predicted by the nearly neutral theory (Ohta, 1973; Zhang, 2003). The ruff, however, has a large historical N_e (>10,000) as determined by genetic methods (Verkuil, 2010), and we therefore exclude the possibility that the gene duplication became fixed due to a recent population bottleneck or low effective population size (see also Lynch et al. (2001)).

In conclusion, the sequence similarity in Domain II can be either the consequence of a process of homogenization or a consequence of recent duplication. We did not compare orthologous and paralogous copies of tRNAs, $tRNA^{Pro}$ and $tRNA^{Glu}$, however, the observed sequence similarity between copies of the tRNAs, $tRNA^{Pro}$ and $tRNA^{Glu}$ indicates that also these duplicate genes are either recent or homogenized (Jiang et al., 2007).

This paper showed that identical gene re-arrangements can occur independently in unrelated birds (shorebirds, tube-nosed seabirds and spoonbills) (Abbott et al., 2005; Cho et al., 2009). If concerted evolution of paralog copies of the control region would apply to ruffs, than there is no active selection against the duplication of genes in the mitochondrial genome. Maintaining two full control regions allows for gene conversion to act as a repair mechanism to avoid transcription error, which might represent a recombinational force in the animal mtDNA as suggested by Tatarenkov and Avise (2007).

Acknowledgments

We thank Oliver Haddrath for inspiring ideas and assistance in the laboratory, and the many insightful discussions. Tissue samples of ruffs outside The Netherlands were collected by Natalia Karlionova, David Lank and Michael Soloviev and other volunteers. Katherina Thuman and Fredrik Widemo kindly provided DNA samples from the breeding population in Gotland, Sweden. Joop Jukema and the Frisian Wilsterflappers traditionally catch ruffs on their Frisian stopovers site and provided us with the chance to collect blood samples and colour-mark individuals which were resighted on breeding grounds in Norway and Siberia. Jos Hooijmeijer provided logistical management of the ruff field work and database. YIV was supported by Dutch Science Foundation (NWO) travel grants (R84-606, R87-307). A.J.B. was supported by an operating grant (200-07) from the Natural Sciences and Engineering Research Council of Canada. This study was financed by a start-up grant to TP from the University of Groningen. Sergio Pereira, Erika Tavares, Robert M. Zink, Ritsert C. Jansen and Irene Tieleman provided valuable advice and insights.

References

Abbott, C.L., Double, M.C., Trueman, J.W.H., Robinson, A., Cockburn, A., 2005. An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in *Thalassarche* albatrosses. Mol. Ecol. 14, 3605–3613.

Avise, J., 2000. Phylogeography. Harvard University Press, Cambridge, Massachusetts.

Baker, A.J., Pereira, S.L., Paton, T.A., 2007. Phylogenetic relationships and divergence times of Charadriiformes genera: multigene evidence for the Cretaceous origin of at least 14 clades of shorebirds. Biol. Lett. 3, 205–209.

Bensch, S., Härlid, A., 2000. Mitochondrial genomic rearrangements in songbirds. Mol. Biol. Evol. 17, 107–113.

Buehler, D.M., Baker, A.J., 2003. Characterization of the red knot (*Calidris canutus*) mitochondrial control region. Genome 46, 565–572.

Buehler, D.M., Baker, A.J., 2005. Population divergence time and historical demography in red knots and dunlins. Condor 107, 497–513.

Buehler, D.M., Baker, A.J., Piersma, T., 2006. Reconstructing palaeoflyways of the late Pleistocene and early Holocene red knot *Calidris canutus*. Ardea 94, 485–498.

Cadahia, L., Pinsker, W., Negro, J.J., Pavlicev, M., Urios, V., Haring, E., 2009. Repeated sequence homogenization between the control and pseudo-control regions in the mitochondrial genomes of the subfamily Aquilinae. J. Exp. Zool. B – Mol. Develop. Evol. 312B, 171–185.

Cho, H.J., Eda, M., Nishida, S., Yasukochi, Y., Chong, J.R., Koike, H., 2009. Tandem duplication of mitochondrial DNA in the black-faced spoonbill, *Platalea minor*. Genes Genet. Syst. 84, 297–305.

Desjardins, P., Morais, R., 1990. Sequence and gene organization of the chicken mitochondrial genome – a novel gene order in higher vertebrates. J. Mol. Biol. 212, 599–634.

Desjardins, P., Morais, R., 1991. Nucleotide-sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. J. Mol. Evol. 32, 153–161.

Eberhard, J.R., Wright, T.F., Bermingham, E., 2001. Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona*. Mol. Biol. Evol. 18, 1330–1342.

Fujita, M.K., Boore, J.L., Moritz, C., 2007. Multiple origins and rapid evolution of duplicated mitochondrial genes in parthenogenetic geckos (*Heteronotia binoei*; Squamata, Gekkonidae). Mol. Biol. Evol. 24, 2775–2786.

Gibb, G.C., Kardailsky, O., Kimball, R.T., Braun, E.L., Penny, D., 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Mol. Biol. Evol. 24, 269–280.

Haring, E., Riesing, M.J., Pinsker, W., Gamauf, A., 1999. Evolution of a pseudo-control region in the mitochondrial genome of Palearctic buzzards (genus *Buteo*). J. Zool. Syst. Evol. Res. 37, 185–194.

Jiang, Z.J., Castoe, T., Austin, C., Burbrink, F., Herron, M., McGuire, J., Parkinson, C., Pollock, D., 2007. Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. BMC Evol. Biol. 7, 123.

Karlionova, N., Pinchuk, P., Meissner, W., Verkuil, Y.I., 2007. Biometrics of ruffs (*Philomachus pugnax*) migrating in spring through southern Belarus with special emphasis on occurrence of faeders. Ringing Migrat. 23, 134–140.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings Bioinform. 5, 150–163.

Kurabayashi, A., Sumida, M., Yonekawa, H., Glaw, F., Vences, M., Hasegawa, M., 2008. Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in mantellid frogs from Madagascar. Mol. Biol. Evol. 25, 874–891.

Lowe, T.M., Eddy, S.R., 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucl. Acids Res. 25, 955–964.

Lynch, M., O'Hely, M., Walsh, B., Force, A., 2001. The probability of preservation of a newly arisen gene duplicate. Genetics 159, 1789–1804.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., 1998. Multiple independent origins of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. USA 95, 10693–10697.

Ohta, T., 1973. Slightly deleterious mutant substitutions in evolution. Nature 246, 96–98.

Paton, T., Haddrath, O., Baker, A.J., 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proc. R. Soc. London, B 269, 839–846.

Pereira, S.L., Baker, A.J., 2004. Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics. BMC Evol. Biol. 4, 17.

Pereira, S.L., 2000. Mitochondrial genome organization and vertebrate phylogenetics. Genet. Mol. Biol. 23, 745–752.

Pereira, S.L., Grau, E.T., Wajntal, A., 2004. Molecular architecture and rates of DNA substitutions of the mitochondrial control region of cracid birds. Genome 47, 535–545.

Quinn, T.W., Wilson, A.C., 1993. Sequence evolution in and around the mitochondrial control region in birds. J. Mol. Evol. 37, 417–425.

Segre, A., Richmond, R.C., Wiley, R.H., 1970. Isozyme polymorphism in ruff (Aves, *Philomachus pugnax*) – a species with polymorphic plumage. Comp. Biochem. Physiol. 36, 589–592.

Singh, T.R., Shneor, O., Huchon, D., 2008. Bird mitochondrial gene order: insight from 3 warbler mitochondrial genomes. Mol. Biol. Evol. 25, 475–477.

Tatarenkov, A., Avise, J.C., 2007. Rapid concerted evolution in animal mitochondrial DNA. Proc. R. Soc. London, B 274, 1795–1798.

Verkuil, Y.I., 2010. The ephemeral shorebird: population history of ruffs. PhD thesis, University of Groningen, Groningen.

Wenink, P.W., Baker, A.J., Tilanus, M.G.J., 1993. Hypervariable-control-region sequences reveal global population structuring in a long-distance migrant shorebird, the dunlin (*Calidris alpina*). Proc. Natl. Acad. Sci. USA 90, 94–98.

Wenink, P.W., Baker, A.J., Tilanus, M.G.J., 1994. Mitochondrial control-region sequences in two shorebird species, the turnstone and the dunlin, and their utility in population genetic studies. Mol. Biol. Evol. 11, 22–31.

Zhang, J., 2003. Evolution by gene duplication: an update. Trends Ecol. Evol. 18, 292–298.